From Game Design to Goal Delineation in the Tandem Transformational Game Design Framework

Safinah Ali[1], Alexandra To[2], Zhen Bai[2], Jarrek Holmes[2], Elaine Fath[3],

Jessica Hammer[2], Geoff Kaufman[2]

MIT Media Lab[1], Carnegie Mellon University[2], Schell Games[3]

Abstract

Transformational game design requires clear delineation of transformational goals, both to guide the game design process and to evaluate a game's intended impact. Tandem Transformational Game Design framework supports game designers and researchers in selecting relevant theories and translating those theories to design decisions using a goal delineation process, which occurs in tandem with the iterative development and playtesting of game artifacts in a game design process. An alignment stage between the goal delineation and game design processes helps multidisciplinary design teams come to common understanding and supports theory-driven prototyping and testing. Previous work has shown how theory development can directly inform the iterative game design process. In this paper, we demonstrate the importance of moving from the game design iteration loop back to the goal delineation loop, using playtest findings. We use our work on the game *Outbreak* to demonstrate how this alignment process might happen in practice.

Keywords : emotional relevance; game design; goal delineation; playtesting; prototyping; multidisciplinarity; transformational games

Introduction

Transformational games are designed with the specific intention of changing players' behaviors, attitudes, or knowledge during and after play (Culyba, 2015). In transformational or educational game design, developing a clear, shared vision of how the player should change as a result of the game is a critical and ongoing process. Given the multidisciplinary interest and expertise involved in designing transformational games, design teams tend to be comprised of members from a broad range of disciplines, each bringing different perspectives, vocabularies, and areas of expertise to the table. Multidisciplinary teams, particularly those comprised of both expert and novice designers and researchers, may experience barriers to defining and sharing their vision due to disparate vocabularies and theoretical frameworks. This can make achieving such unification of vision quite challenging. Adding to a growing body of research that attempts to tackle this challenge, the Tandem Transformation Game Design framework proposes the two mutually informing "loops" of theory-driven goal delineation and goal-driven game design. Previous work from our team has outlined the overall framework, and demonstrated through a design case study how the goal delineation loop informs the game design loop. In this paper, we discuss how insights and observations from the game design loop can often necessitate a return to the goal delineation loop through an alignment process that can inspire the team to reconsider and refine the key theories informing the designation of a game's core transformational goals and the mechanisms for achieving them.

To understand this process, we first review the Tandem Transformational Game Design Process. Then we talk about how, in an attempt to design game-based interventions for fostering curiosity through play, we reviewed literatures of curiosity from psychology and came to an initial understanding of the theoretical space and selection of focal curiosity-related constructs to inform the designation of the game's specific transformational goals. We detail how this understanding of theory in turn informed our game iteration loop while designing the game *Outbreak*. We then outline what we learned through playtesting of early versions of the game, and how playtesting data informed not only the design of the game, but also inspired us to return to the literature to achieve a deeper understanding of the theoretical space and a more nuanced delineation of the game's intended goals. Specifically, based on our observations and measurements of players' emotional responses to *Outbreak*, we returned to the psychological literature on emotion both to help make sense of what the playtesting revealed and, moreover, to begin to apply those theories to iterative design decisions. In sum, in this paper, we focus on the importance of returning from the Goal-Driven Game Design loop back to the Game-Driven Goal Delineation loop, and explain how our team navigated the alignment between these two loops. This paper thus aims to provide, through an illustrative case study, a design framework resource for multidisciplinary transformational design teams to employ in their own practice.

Tandem Transformational Game Design Processes

Culyba (2015) defines transformational games as games designed for facilitating how the player is changed outside the game. While game designers are proficient at creating fun and engaging games, transformational games require designers to consider psychological and social factors that affect players' willingness and ability to change (Culyba, 2015), and to have a deep understanding of relevant content. Making a viable, entertaining, and effective transformational game means drawing on theories and methodologies from a range of fields in addition to game design, such as psychology, learning sciences, and human-computer interaction, and deciding how to integrate or translate those theories and methodologies to specific game design mechanisms and content (Seidman et al., 2015).

In order to draw from such diverse areas of theory while designing effective games, transformational game design teams often bring game designers together with experts in the design and research of psychological or educational interventions. Helping these interdisciplinary teams collaborate effectively is challenging but critical. One of the biggest predictors of a game's success is the team's level of clarity and alignment on the vision (Tozour, 2015). Moreover, given the potential for players' resistance to overly didactic or prescriptive game experiences, the overt insertion of existing intervention methods or strategies into game contexts can reduce how engaging and/or impactful a transformational game is likely to be (Kaufman & Flanagan, 2016). To this end, game design approaches that aim to more fully integrate or embed known theories or interventions in the design process tend to produce better results and, from players' perspective, better games. In order to accomplish this successful intermingling of disciplines, iterative, player-centric game design methods such as rapid prototyping must be combined with psychological insights (Seidman et al., 2015; Flanagan et al., 2013). To this end, psychologists (and other domain experts) must join forces with game designers, who, domain experts in their own right, often lack firmly established ways to bring non-game designer team members to the table, and vice versa. Non-designer content experts, meanwhile, may struggle with translating their knowledge of abstract concepts and principles from a given theoretical literature to concrete game design decisions. While transformational game design teams do, in practice, find ways to build on these multiple intellectual traditions, the integration is often challenging.

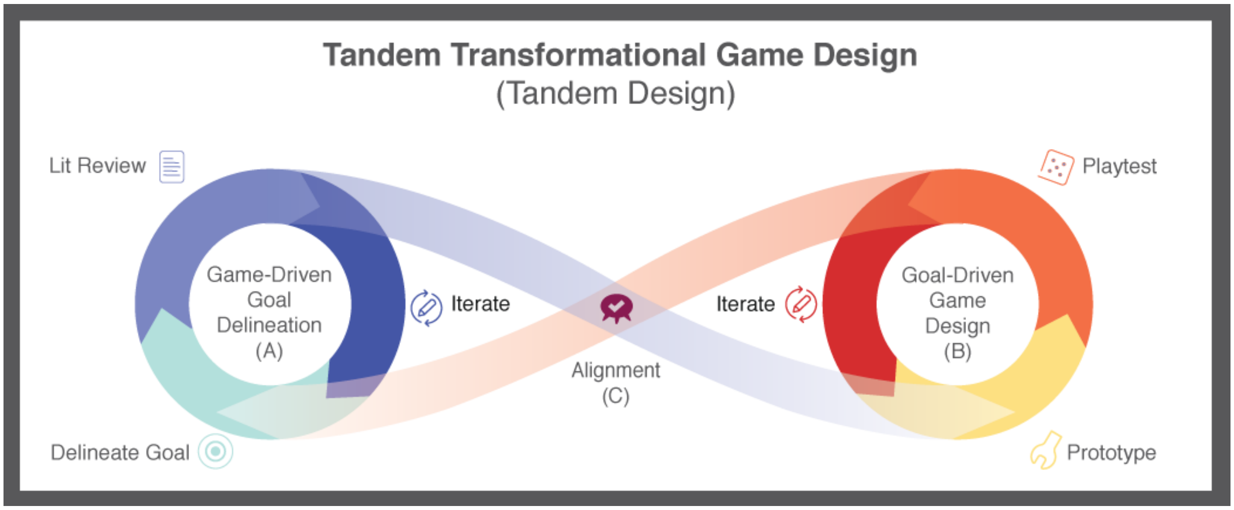


Figure 1. The Tandem Design Process. **Delineate Goal:** articulation of goals for player transformation. **Lit Review:** collective reading of prior research to develop a shared vocabulary. **Alignment:** The juncture between the two phases. See descriptive paragraph below. **Prototype:** rapid game prototyping with one or more *delineated goals* as the initial point of inspiration. **Playtest:** frequent and extensive testing and evaluation of game (both within the team and with target audience). **Iterate:** the process of refinement when remaining inside one cycle. *Icons from (Harlow, n.d.; Luck, n.d.)

The Tandem Transformational Game Design Framework (Figure 1) is one method that addresses this challenge. In the Tandem Design process, prototyping games and articulating goals are positioned as intrinsically intertwined cycles occurring in tandem with one another. In the *game-driven goal delineation* cycle (goal cycle, Figure 1.A), teams articulate their goals in conversation with the research literature and other data sources; in the *goal-driven game design* cycle (game cycle, Figure 1.B), teams prototype and playtest their games. Games and other research artifacts are used to iteratively align a team's prototyping process with their vision and goals (alignment, Figure 1.C). Prior work has shown how theory can guide the development of game prototypes and inform specific game design decisions (Figure 1.A). The current paper examines how prototypes and playtest findings can, through an alignment process, inform the iterative development of the theoretical foundations informing the team's game design goals (Figure 1.B, Figure 1.C).

In this case study, we focus on a game-based intervention for fostering curiosity through play, one aiming to increase young players' comfort and engagement with science, technology, engineering, and math (STEM) topics. During the design process, the team used the goal cycle of Tandem Design to develop a better shared understanding of the *problem space* - the conceptualization and operationalization of the construct of curiosity; the *intended audienc*e - underrepresented groups in STEM, including minority, women, and low socioeconomic status students; and the *transformational goals* - increasing curiosity through play (To et al., 2016). In prior work on the design process for this game, the team described researching *theories* of curiosity and developing *elements* of curiosity that could be instantiated or explored in a game (To et al., 2016). They developed *artifacts* to use during the alignment process, such as lists of game elements (Figure 3), and used those artifacts in the alignment process to decide whether they were ready to move to the prototyping stage or whether they needed to iterate through the goal delineation cycle (Figure 2) again.

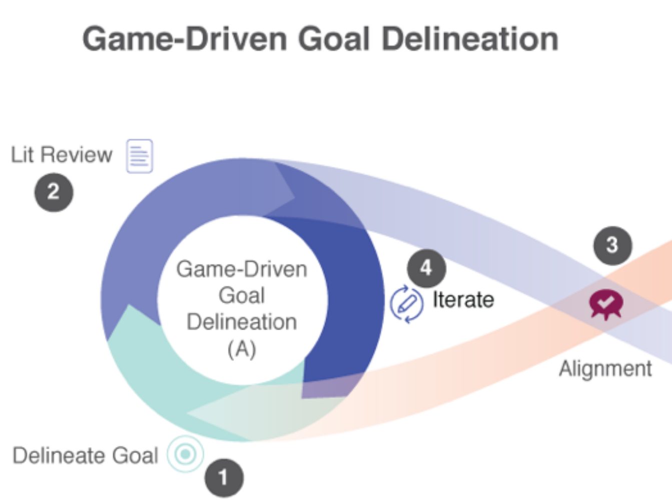
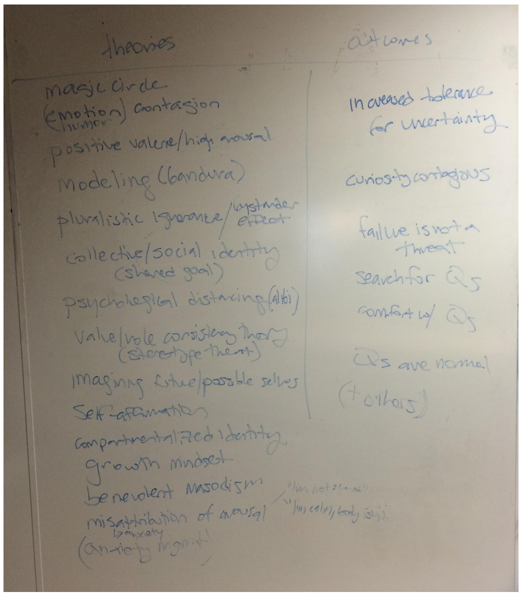


Figure 2. Our team’s pathway through our first game-driven goal delineation cycle.



Theories	Outcomes
Magic circle Emotion Contagion Positive Valence/High Arousal Modeling Pluralistic Ignorance Collective/Social Identity Psychological Distancing Value/Role Consistency Imagining future and past selves Self-affirmation Compartmentalization of identity Growth mindset Benevolent Masochism Misattribution of Arousal	Increased tolerance for uncertainty Curiosity contagion Failure not a threat Search for unanswered questions Questions are Normal

Figure 3. Documentation photo (left); transcription (right). After reviewing the curiosity literature, the team extracted the elements (right, then labeled outcomes), brainstormed related moments in their own lives out loud, and extracted related theories from those stories and goals.

After a second pass through the goal cycle, the team moved to the goal driven game design loop (Figure 4). They documented their prototyping and game development process (To et al., 2016), building on existing design practices such as reflective practice, rapid prototyping, and iterative design (Zimmerman et al., 2014; Frens et al., 2013; Nielsen 1993;) and integrating them with current game design practices such as playtests. The team ideated a wide range of game ideas (using the theories and intended outcomes listed in Figure 3 as a foundation), created playable prototypes to use as internal reflective probes, iterated on the prototypes to create designs for the target audience, and playtested the games repeatedly. They implemented design practices such as parallel ideation and removal of individual ownership of game concepts in order to develop the widest range of interesting and successful ideas.

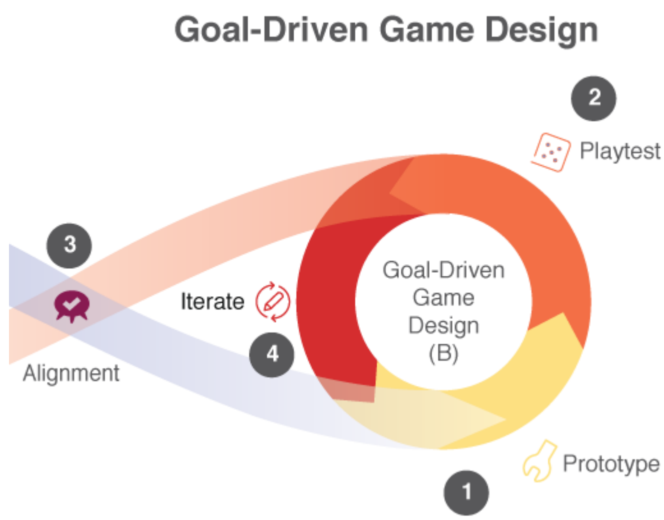


Figure 4. The Goal-Driven Game Design Cycle. Now with delineated goals, team members divergently prototype (1), playtest within the team (2), align with the team on what's working (3), then either move on to game iteration (4) or revisit the goal cycle.

In the end, *Outbreak* emerged as one design that the team brought to fruition, and elsewhere we have detailed the initial processes of goal delineation and game design that guided its development (To et al., 2016). However, the Tandem Design process does not stop when a game has been successfully created. The prototyping and playtesting process ends by returning to the alignment process (Figure 1.C) and is meant to inform the iterative development of further theory (Figure 1.A) and/or the redesign of a game in light of new theoretical insights and frameworks that emerge. It is this process that we explore in depth in the present work. Following a brief overview of the game's design, we discuss how the results of our game's playtesting (in particular, the emotional responses evoked by the game) inspired a return to the literature to help make sense of these results and to refine the transformational goals of the game. In the following section, we outline the design decisions of the game in question, *Outbreak*.

Outbreak

Outbreak is a cooperative tabletop game for two to five players, targeted toward ages 9-14, in which the group must save a town from a rogue scientist by searching their laboratory for antidotes to a disease, and asking questions to solve puzzles. Most players assume the role of scientific investigators, but one player takes the role of the investigators' robot assistant. Each time the investigators encounter a new room in the laboratory, they can send the robot in first to determine what potential dangers or obstacles await (e.g., locked cabinets, unfriendly creatures) and what character traits (e.g., speed or stealth) and resource cards (e.g., flashlight or lockpick) might be helpful in overcoming them. Investigators then enter the *question-asking phase*, where they can ask the robot questions to learn more about the challenge they will face in the room and prepare accordingly; however, the robot can only answer questions that are posed to it using special formats that the robot can understand (represented as “question tokens”). After the *question-asking phase* of the game, players enter the *discussion phase*, during which the players must collaboratively decide what resources to use in order to overcome the room’s challenges and claim its antidotes. If the players are successful, they receive antidotes; if not, they lose the resources they used to make the attempt.

Learning from literature, we structured the design goals behind *Outbreak* to operationalize curiosity through two specific curiosity elements: (1) *comfort with uncertainty* which relates to players’ perceptions of failure, their comfort and willingness to take risks, and their search for unanswered questions and (2) *comfort with questions*, which relates to players’ perceived abilities to fill a knowledge gap and cope with uncertainty, their persistence towards understanding, and their assessment of their own knowledge states. *Outbreak* was designed as a multiplayer board game, which necessitates that players are co-present during gameplay, in order to help mitigate the affective experience of failure and the strategies available to manage it (Mohammed & Dumville 2001). Specifically, the question-asking phase was designed to encourage equal and frequent participation from players and to foster a sense of a shared experience of uncertainty (Mahdikhani et al. 2015). Second, the game supports risk-taking and failure by allowing players to experiment with different hypotheses and, together, to confront the results of incorrect assumptions or suboptimal decisions (Rocca 2010).

In the following sections, we discuss our methods and processes of playtesting *Outbreak* and what we learned from the playtest results.

Summary of Playtest Methods and Results

In order to examine the emotional impact of playing *Outbreak*, we conducted playtest sessions of the game in a lab setting, while we recorded and analyzed pre- and post-play self-report measures of emotion as well as gameplay behavior. We recruited 21 local Pittsburgh participants (9-14 y/o, 9 female) and ran hour-long playtest sessions with a total of 10 groups (each comprised of 3-4 participants).

We recruited participants by distributing advertisements via local public schools, parent mailing lists within the university's research participant pool, local parent groups on social media (e.g., Facebook, Craigslist), as well as local community centers. Participants and their parents/guardians were provided with a consent form explaining the study's purpose, procedures, participant requirements, potential risks and benefits, and confidentiality assurance, as well as a request for permission for audio and video recording, demonstration and online crowdsourcing and third party services for transcription. All participants' parents gave consent, and participants provided assent by signing the consent form on the day of study.

Participants were seated around a square table in a private lab space, with a researcher present at all times to facilitate game play and distribute study materials. First, participants were equipped with audio recording collar microphones and then performed a number of lightweight pre-playtest icebreaker activities - the results of which were not significant to the findings of this paper and will be reported in future work.

Next, participants played *Outbreak*. The researcher initiated game play by explaining the rules to the participants and then running a scripted practice round with the participants to demonstrate how to play. As a part of the script, the researcher directly encouraged players to provide their honest feedback about the game when appropriate and disclosed that they did not participate in the design of the game. During play, the researcher assumed the role of “robot,” with participants playing the role of investigators. The game challenges were randomized for each play group, but the researcher was trained with guidelines for responding to players’ questions to increase uniformity between groups. Game play then proceeded either for 40 minutes or until the participants reached the end of the game.

Measures

In our playtest we recorded and analyzed game play behavior data as well as pre- and post-play measures related to player curiosity and players' affective experiences. During game play, video and audio data were recorded. Additionally, a researcher seated in the room recorded structured field notes, specifically focusing on inter-player dynamics. Quantitative field note data included the number of questions asked, challenge successfully completed, conflicts, and agreements. Qualitative field note data included instances of emotional expressions (e.g., utterances such as "yay!", or "I don't think we can ever win" and occurrences of smiling or frowning) and specific actions (e.g., silence for an entire round, moderating discussion, grabbing other players' cards).

In addition, in order to examine how participants responded emotionally to different phases of game play, we asked participants to complete a self-report measure of emotional experience. This measure specifically gauged the levels of valence and arousal experienced by players during or after different events in the game and was administered individually on a laptop computer. Participants were provided with a four-quadrant map on the screen with different emotions at each edge and corner of the map corresponding to different combinations of positive/negative valence and low/high arousal. The interface displayed 15 prompts for stages of the game (e.g., `When you won a round', `When someone disagreed with you') , and the participants were asked to click the point on the map that best corresponded to how they felt during that stage of the game. Events were chosen in order to provide more fine-tuned insight for understanding when and how players experience different emotions during the game. This method of inferring valence and arousal of the participants provided us more detailed quantitative data compared to previous work, where we qualitatively coded these events along a variety of axes - most relevant to this work we include codes for whether or not the event is a "positive, neutral, negative, or unclear" event (where an event like winning a round is positive and an event like losing a resource is negative while discussion is neutral).

Coordinate (x, y) data, corresponding to the click location of the valence/arousal map on a -2.5 to 2.5 scale, were recorded for each participant. In our analysis for valence on the x-axis we defined the coordinate range from -2.5 to -.5 as 'negative' affect, -.5 to .5 as 'neutral' affect, and .5 to 2.5 as 'positive' affect.

In the analysis we aggregated the valence/arousal data across all study groups and examined the valence response (DV) by question-prompt (IV) or game phase (IV) as well as the relationship between valence (DV) and arousal (IV).

Two prompts were related to question-asking phase of the game :

- 'The first round of asking questions to the robot',
- 'The last round of asking questions to the robot.'

We use these two events in order to look for overall change in emotion towards question-asking across a game session.

One prompt related to the discussion phase of the game:

- 'Discussing which room to go into.'

Other prompts that were related to question-asking and discussion had clear emotional biases, and hence, were omitted. For example, ‘When someone disagreed with me’ is an event that might happen during discussion, but is more likely to be associated with negative emotions. We analyze the average response valence for these three prompts to indicate the emotions that participants felt during those two phases of the game.

In order to examine sentiment and emotion expressed in players’ dialogue during the game, we used IBM Watson's sentiment analysis tool (High, 2012) on transcripts from the audio recording in order to examine what emotions players verbally expressed during play.

From Artifact to Theory

In this section of the paper we will detail the alignment process that occurred after several cycles of iterative game design as an illustrative case study for how a team might use playtest data to move from the game design and iteration loop of Tandem Design back to the goal delineation loop.

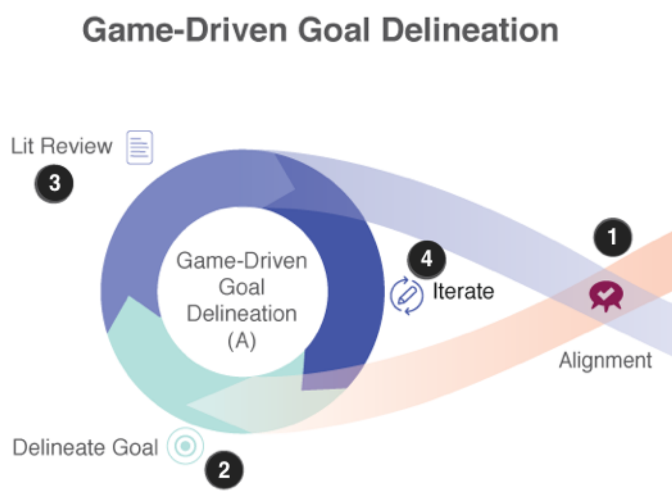


Figure 5. The Goal-Driven Game Design Cycle. Now with delineated goals, team members divergently prototype (1) alignment of the team around the players' experiences and responses, (2) Goal Delineation process, (3) Review emotion theory literature, and (4) Iterating on goal delineation to return to common alignment of the team.

As academic researchers, our alignment process at this stage was centered around organizing, analyzing, and interpreting the playtest data, both individually and during large and small group meetings, with the initial goal of publishing the insights we gleaned regarding players' experiences and responses to the game. We understand this process as alignment, as it involves the entire team coming to a central understanding of the project (Figure 5.1). At this stage, we analyzed the playtest results and came to a common understanding of the players' emotional responses in the game. This transitioned into the goal delineation loop that was based on a new theoretical framework that ties in the theories of emotion that we reviewed to explain our playtest results (Figure 5.3). We outline this process in greater detail below.

In order to examine sentiment and emotion expressed in players' dialogue during the game, we used IBM Watson's sentiment analysis tool (High, 2012) on transcripts from the audio recording in order to examine what emotions players verbally expressed during play. The tool takes in transcripts as text files and returns a numerical result for the fear, disgust, joy, sadness, anger, and overall sentiment expressed in that text data using deep natural language processing (i.e., gathering and applying as much relevant context from the text as possible). We segmented game transcripts and extracted text related to the two main phases of game play: 1) question-asking and 2) discussion. Researchers who were not involved in running the study transcribed the audio and denoted the beginning and ending of the question-asking and discussion phases using verbal cues from the robot player (e.g., "you can ask questions now", "ok time's up on question-asking"). We then performed IBM Watson analysis on transcript data from each category separately. Taking the results from the IBM Watson analysis, we compare the aggregate question-asking phase analysis to that of the aggregate discussion phase analysis.

From this sentiment analysis we used confidence intervals to examine the difference between these phases of game play and found that players expressed significantly more negative overall sentiment, more disgust, and more fear during question-asking than during the discussion phase of game play. There was no significant difference in expressed joy, anger, or sadness. This differed from self-report data in which we saw that participants reported positive and neutral affect during both question-asking and discussion.

This divergence between the detected and self-reported emotions experienced by players during these key game rounds was unexpected - and, as we discuss below, inspired us to return to the psychological literature on emotion in order to try to make sense of these findings. In our final round of alignment, we thus looked closely at this gap between expressed and self-reported emotion and called upon various theories of emotion to offer potential clarification. We considered several possible explanations for the difference in participants' self-reported and observed emotional experiences during play. To better understand why the players reported their emotion differently, we ventured into the domain of affect and emotion, and expression of emotion, and explored several theories that help us make sense of our results. We found three strong contenders for what might explain this phenomenon: misattribution of arousal, benign masochism, and social facilitation theory. This new literature review, which we outline in the next section, transformed the team's understanding of the theoretical space.

Theories of emotional relevance to situation

As a psychological, behavioral and affective state, curiosity is associated with a variety of arousal states, combining the aversive feeling of knowledge deprivation, the pleasure of knowledge acquisition, and the positive anticipation of new information (Litman, 2010, Jirout and Klahr, 2012). Much research suggests that curiosity and anxiety are co-activated in response to approach or avoidance of uncertainty, and there is an optimal threshold between curiosity and anxiety (Kashdan and Roberts, 2006). One explanation of the negative correlation between curiosity and anxiety is that the latter increases a preference for low-risk options, which prevents people from approaching uncertain scenario (Lerner et al., 2015).

The design of *Outbreak* aimed to incorporate opportunities for players to manage negative emotions around question-asking through play. This paper's key finding, that players expressed negative sentiments during game play but reported positive emotional reactions when looking back on their experiences in the game, suggests that players are indeed successfully managing or shifting their negative emotions. This specific pattern of results is consistent with a range of psychological constructs and theories pertaining to the relationship between arousal and emotion, including misattribution of arousal, benign masochism, and social facilitation theory.

We consider the role each of these explanations plays in helping to account for the difference between participants' observed and self-reported emotional experiences of play.

Misattribution of Arousal

Cognitive reappraisal is a main emotion regulation approach that helps shift negative emotional states to more positive ones through reinterpreting the meaning of anxiety-inducing situation (Gross and John, 2003). Common reappraisal approaches involves eliminating negative arousal by helping people recognize that there is minimal threat or negative consequences (Hofmann, 2009), or enhancing positive arousal by emphasizing the pleasurable outcomes involved in the situation (Fredrickson, 2001). Previous research has demonstrated promising emotion regulation effects in fostering curiosity for young children by using intelligent learning companions to reduce anxiety for potential failure (e.g. ""I love getting it wrong sometimes. This is how you learn new things") and elicit positive emotion relating to knowledge acquisition (e.g. "I love to learn") (Gordon, 2015). Cognitive reappraisal involves a relatively high cognitive demand, as it requires a two-step process: the subject has to first attribute causations of one's own emotional arousal, and then actively modify the meaning of the associated consequences. An alternative approach to cognitive reappraisal is to influence at the first step by changing the interpretation of the attribution. For example, research shows that by shifting the attribution of high arousal from anxiety to excitement based on their arousal congruency, people perform better in anxiety-inducing activities such as public speaking and math tasks (Brooks, 2014). Research also shows that by misattributing the source of arousal, unwanted effects of emotion on decision making can be reduced. There is an explicit link between behaviors and reported emotional states. In studies describing this phenomenon there is an observed difference between subjective reported experience of emotion and an objective measure of emotion. For example people may report: less situational anger when they can attribute negative feelings to drug-induced arousal, higher confidence in their ability to achieve tasks when they can attribute anxiety and nervousness to subliminal noise, and less error-related negative emotions when completing error-prone tasks when they could attribute aversive feelings towards a placebo pill.

As predicted by prior work, the challenges of question-asking (e.g., its perception as a socially risky, anxiety-inducing activity) may have produced heightened levels of arousal for players during the question-asking phase of the game. However, the game itself was designed to offer a number of arousal-inducing elements that could serve as plausible explanations for that heightened arousal. Our prior work showed, for example, that players are immediately well-attuned to the scary aspects of the game's theme and narrative. In fact, in early iterations of the game we found that if the arousal and/or anxiety induced by game elements (e.g., the use of scary music) was too high, it interfered with the success of the game's other mechanics to create a safe space to ask questions. For this reason, we believe that these competing explanations with high arousal congruency offered an opportunity for players to interpret their arousal in ways that were less threatening to their self-image and more conducive to their continued immersion in and enjoyment of the game experience. Indeed, prior work has demonstrated that when serious or learning activities are framed as a game, participants tend to view the activity more favorably (Lieberoth, 2015) - after all, games are "supposed to be" fun. Thus, after play, participants, as a result of misattributing the source of their arousal to the game's aesthetics, reflected on their experience of question-asking in a more positive light.

Benign Masochism

Relatedly, benign masochism describes the phenomenon where pleasure is derived from the realization that the brain has falsely interpreted an initially negative experience as threatening

(Rozin et al., 2013). Examples of this phenomenon include eating spicy foods, riding roller coasters, and engaging with horror or suspense narratives. Benign masochism provides a framing for how and why players would enjoy or describe enjoying playing games with a haunted house theme. According to this viewpoint, players' heightened anxiety during the question-asking phase could have signaled to the brain that a threat was imminent, but, upon reflection, the safety of emerging from the experience produced pleasure. This pattern of "hedonic reversal" has been demonstrated in a number of related domains, including the enjoyment of horror films and haunted houses, in which the relief of escape and the recognition of safety can produce intense pleasure. Similar intensification of joyous emotion has been found due to leftover of excitation from previously experienced aversion feelings according to the excitation-transfer theory (Bryant, 2003). This interpretation of the findings assumes that the shift from negative to positive affect occurred at the end of the game, when players' "safety" was fully realized. This is arguably borne out by the finding that there was not a significant difference in sensed (negative) sentiment levels between the first and last rounds of question-asking.

Social Facilitation Theory

Finally, social facilitation theory posits that the mere presence of others heightens physiological arousal, which in turn increases the likelihood of individuals exhibiting their dominant response (i.e., the response that is the most automatic, well-learned, or likely to be the default for that context) (Zajonc, 1965). For example, in the presence of an audience expert pool players were shown to perform better (i.e., more accurate, better shots) while novices performed worse (i.e., less accurate, fewer shots). This theory suggests that players would experience a heightened level of arousal when playing a multi-player cooperative game (as compared to a solo game). This phenomenon may contribute to overall heightened levels of arousal that players might experience, thus supporting players in responding to the other built-in arousal-heightening elements and mechanisms of *Outbreak*. Social facilitation might also then help predispose participants towards emotional re-framing, misdirection, and reinterpretation. However, social facilitation theory predicts that players should experience similar levels of arousal during the question-asking and discussion phases of the game, since both involve players working in a group. This pattern was not observed in our data.

Discussion

Outbreak and theories of arousal

The design of *Outbreak* aimed to incorporate opportunities for players to manage negative emotions around question-asking through play. Through iterative play testing, we observed that players expressed negative sentiments during game play but reported positive emotional reactions when looking back on their experiences in the game, suggesting that players are indeed successfully managing or shifting their negative emotions. This specific pattern of results is consistent with a range of psychological constructs and theories pertaining to the relationship between arousal and emotion, including benign masochism, social facilitation theory, and misattribution of arousal. Interpreted through the lens of benign masochism theory, the discrepancy between the participants' observed and self-reported emotional experiences experiences observed during game play and those reported by players post-game could reflect the affective shift that occurs when an initially negative experience, but, upon reflection, the safety of emerging from the experience produced pleasure. Findings could also be explained by referring to the social facilitation theory, that suggests that players would experience overall heightened levels of arousal when playing a multi-player cooperative game, which may in turn contribute to supporting players in responding to the other built-in arousal-heightening elements of *Outbreak*. Findings could also be explained by the phenomenon of misattribution of arousal. As predicted by prior work, both the challenges of question-asking (e.g., its perception as a socially risky, anxiety-inducing activity), and the design elements of the game could serve as plausible explanations for that heightened arousal. Thus, after play, participants, as a result of misattributing the source of their arousal to the game's aesthetics, reflected on their experience of question-asking in a more positive light.

Of course, without direct evidence of players' cognitive appraisals of their emotions, it becomes difficult to disentangle these theoretical explanations. While both benign masochism and misattribution of arousal involve key processes involved in interpreting the meaning of their feelings, they diverge in how the reconciliation between negative and positive emotions occurs. Benign masochism predicts a shift from negative to positive emotional experiences as the transition from perceived danger to safety occurs. In contrast, misattribution of arousal suggests that the shift happens not at the emotional level but at the level of interpretation preceding the experience of emotion (that is, the appraisal of the meaning and valence of one's felt arousal). In other words, benign masochism entails ambivalence (in the co-occurrence or shift between negative and positive emotions), whereas misattribution of arousal entails ambiguity (with the valence of experienced emotions dependent on how the brain interprets the cause of arousal).

The difficulty of differentiating between these two mechanisms as explanations of player emotion, and the potential discrepancy of real-time rating of “experiencing self” and retrospective report of “remembering self” due to bias of peak and end feelings (Kahneman, 2003), point to the need for additional measures, such as players' cognitive appraisals and real-time reports of affective experience, in future work. What we hope to have demonstrated is that games can indeed by designed to facilitate the manipulation or re-direction of player emotion, especially in ways that shift perceptions of content or behaviors that may elicit negative emotions when encountered outside the context of the game.

Finally we observe that for positively coded events, the higher the arousal the more positive the valence and for negatively coded events, the higher the arousal the more negative the valence. This may indicate that participants are "correctly" attributing their arousal based on the valence of the game event. That is, this is in line with appraisal theories of emotion that suggest we use the context to interpret the meaning of our arousal, which leads to emotional experience in line with the valence of the contextual cue (Scherer, 2001). This may also align with our hypothesis that players are misattributing their arousal during play, although in this case they may be using contextual cues rather than game elements to make emotional attributions.

Process insights

In our prior work on Tandem Design we stressed the dual importance of defining transformational goals and producing game artifacts, presented a method for cycling between those two phases of game design, and shared a case study focused on game goals (To et al., 2016). The present work emphasizes how working with artifacts and playtest data can generate not just research insights, but new design goals. We accomplish this through the process of alignment, as shown in the case study above.

During our alignment process, we identified strategies that helped our interdisciplinary team accomplish critical tasks: 1) make sense of complex data, 2) retain the connection between artifact design and player data, and 3) translate research hypotheses into transformational design goals. We consider each of these challenges in turn.

Making sense of complex data. Post-playtesting, we had a range of data sources, including observational field notes, transcripts, audio and video, and quantitative measures. In our interdisciplinary design team, members had different levels of expertise in analyzing or interpreting data. Aligning the team members' understanding of the playtest data was therefore a challenge. To address this challenge, we assigned some team members to produce interpretable artifacts for each data source, such as data visualizations of the valence arousal data points. However, because the point of the process was *alignment among team members*, and utilize their range of expertise, we did not want to have the quantitative data, the qualitative data, the design team, etc. operate in a vacuum. We therefore connected the interpretable artifacts to one another. For example, the valence-arousal visualizations were discussed in the context of selected quotes and gameplay moments drawn from the qualitative data. One of the challenges faced while trying to intertwine multiple data artifacts was that different segments of data analysis progressed at different speeds. For instance, quantitative analysts were able to bring in new visualizations every week, while our qualitative coding process was slower. In future iterations of this process, we plan to allot proportional resources and time to keep the analyses concurrent.

Connecting artifact design and player data. In order to understand how to iterate the game, we needed to understand how our design decisions were driving the playtest findings we observed. However, in an iterative playtest process, it is not always possible to formally test every design decision. We needed to rely on game design analysis techniques, such as identifying and coding game events, to build hypotheses about what design decisions might be driving player reactions. While we were able to build on existing game design knowledge in the original design process for *Outbreak*, by connecting theories of curiosity to uncertainty in game design, the second phase of the process did not immediately suggest an existing body of game analysis literature to draw from (To et al., 2016). We addressed this issue by developing custom measures for our own game, such as game event prompts for self report. These events were chosen as significant game moments, and coded as positive, neutral, negative, or unclear, to provide fine tuned insight about what event has an emotional effect on the player, and by what amount. However, this opens the possibility of disproportionately weighing some events and omitting others. One method to analyze emotion data that could have been used is to look at all significant deviations in valence and arousal in the sentiment analysis data points, and include all the game events corresponding to those deviations in the self report, in addition to what team members identified as significant events.

From research hypotheses to design goals. The purpose of this process is not just to develop research insights, but to identify ways to iterate *Outbreak* so that the game better aligns with its transformational goals. As part of the case study presented here, we identified three theories related to emotion that could explain our playtest findings. While there is value in disambiguating between the three theories, we asked to what extent these three theories implied different *design goals* or different *game design decisions*. For instance, while benign masochism predicts a positive emotional shift due to reduced perceived danger, misattribution of arousal suggests that the shift happens at the level of interpretation preceding the emotional experience, as opposed to the actual emotional experience. This helped us acknowledge that the extent to which reduction of perceived danger in the game design has an effect on the participants' emotional shift, is not certain. As part of this process, we also recognized that we could work toward increasing the emotional impact of *Outbreak* even before we disambiguated between the competing explanations for our findings. Our findings were not consistent with a pure social facilitation hypothesis, but they were consistent with both misattribution of arousal and benign masochism. In turn, both those theories suggest that adjusting the frightening elements of the game would affect player behavior. Hence, we are now exploring a redesign of *Outbreak* in which we amplify the scary or stressful game aesthetics to help players cope with emotionally or socially difficult learning tasks. Our model currently assumes that teams are *either* in a goal delineation *or* in a game design phase, and outlines the process our team followed in adjusting the goal, or iterating the game. We have found that sometimes the appropriate outcome for an alignment process is for the team *both* to adjust their goals (e.g. by performing research studies to disambiguate competing hypotheses) *and* to continue designing and iterating games.

Limitations and Future Work

We acknowledge that there may be multiple interacting psychological phenomenon that contribute to the gap between reported and observed affect during play. For example, although we strongly believe that players may be misattributing the source of their arousal, it is difficult to isolate this phenomenon with our current data. In order to understand the role of game theme and content in inducing or appearing to induce physiological arousal, further inquiry is required. For example, we could examine *Outbreak* players' emotional responses not only to game events but to game content. Additionally, in our sentiment analysis, we only examine the question-asking and discussion phases of the game. To explore whether or not participants are experiencing benign masochism we may want to examine post-round and post-game conversation to see how participants react to both the results of their discussion and decision making as well as the results of the game as a whole. This future research would allow us to isolate the psychological mechanism at play, discover ways to amplify it, and more reliably reproduce it in the future.

Nonetheless, there are still immediate implications for game designers who want to help players cope with anxiety-inducing or otherwise emotionally challenging tasks. Rather than avoiding additional potential stressors, designers can leverage scary or stressful game aesthetics to help players cope with emotionally or socially difficult learning tasks. The different psychological theories we consider in this paper imply that the mechanism for this coping behavior might manifest in different ways, but for many designs the two can be treated equivalently.

We recognize limitations in our iterative playtesting and data analysis cycle. Teams can intertwine multiple data artifacts collected during gameplay to better analyze contextual data. Proportional resource allotment to qualitative and quantitative data interpretations can also help collectively analyze large amounts of data better. In identifying significant game events based on designers' discretion, we come across the possibility of being biased, and disproportionately viewing all events. We also assume that teams are either in a game design or a goal deliation stage, but do not consider that the appropriate outcome for an alignment process for a team can be to adjust their goals and to continue designing and iterating games simultaneously.

There are also implications for the iterative design process and playtesting. In light of our findings, how do we interpret players' negative sentiments expressed during gameplay? When do negative sentiments mean the players are having a bad experience, and when does it mean that these alternate psychological mechanisms are being triggered? How does that interact with the ethics of working with children? If players are having a bad experience in the moment, even if they remember it fondly later and it contributes to long-term ability, should the play session continue? One way to approach answering these ethical questions is to develop new methods for analyzing players' emotional reactions. Here the distinction between mechanisms may become more meaningful (e.g., how can we register benign masochism before the player realizes the experience is benign?).

Related to this, in the immediate future we also plan to explore other avenues of research. We will continue to make design changes to *Outbreak*, introducing new sources of arousal through game content, theming, as well as a virtual peer player who can reflect, amplify, encourage, etc. the emotions the players express. We will explore this game design space and seek to identify potential differences in sources of arousal and their effect on misdirecting the source of anxiety as well as question-asking behavior. For instance, one redesign project initiated by the team after reaching this new theoretical understanding, is designing a scary and non-scary version of *Outbreak* and observing the differences in expressed and self-reported emotions, to understand the role of misattribution of arousal.

While misattribution of arousal and benign masochism are both theories about the individual, *Outbreak* is still a group game. It is therefore critical to understand how these mechanisms function at a group level as well as the contributing role of group phenomenon such as social facilitation. Future analysis will focus on both individual and group emotions. One way to explore this question is to develop a virtual peer player who can systematically affect the group-level behavior through interventions such as reflect, amplifying, encouraging, etc. the emotions that players express. We can use the virtual human player to identify the impact of group-level factors and inter-player interactions on players' emotional experiences both during and after play.

Finally, in considering these explanations, we must also rule out possible sources of bias in our measures. When collecting self-report data, there is always the possibility of response bias, and in particular the possibility of reporting more positive emotions in order to please the researchers (e.g., social desirability bias (Paulhus, 1991). In our data collection, we addressed this bias by making it clear that the researchers present were not involved in game design, and sought honest feedback from participants to improve the game. We observed that indeed, participants reported a range of positive and negative emotional experiences across game events, demonstrating that participants are willing to report negative feelings.

Conclusion

This paper describes the team's alignment process after playtesting Outbreak to make sense of findings around players' affective states during and after a transformational game that is designed to facilitate negative emotions such as anxiety aroused by inquisitive learning behaviors, in this case question-asking. We compare players' sentiment analysis of playtesters' in-game verbal responses during game phases with their post-game self-reported emotional data, to draw conclusions about playtesters' direct emotional responses to in-game mechanics against what they interpreted those emotional responses to be attributed to. In particular, we found an intriguing disparity between observed and self-reported sentiment, where players expressed negative sentiments during game play but reported positive emotional reactions on reflection of their in-game experience. These findings are consistent with explanations of misattribution of arousal or benign masochism, both of which deal with players' interpretations of their own feelings based on experiences in the game, but not with social facilitation theory, which is purely an effect of group play. We discuss how, as a team, we came to a common understanding of the playtest results, and how it lead us to explore several theories of emotion and affect. This exploration of emotional theories led us to develop a new shared understanding of the theoretical space, which in our future work, will influence the game design of *Outbreak*.

Additional research is needed to disambiguate benign masochism and misattribution of arousal as the mechanism, and to better understand group-level effects. This paper helps design teams in designing games with elements that promote positive learning by redirecting negative emotions around anxiety-inducing learning activities in group play, as well as make use additional theories to make sense of playtest data and bring a diverse team to a common understanding of the data.

Our playtest findings suggest that playtesters self-report no significant difference in their emotional states in high- and low-uncertainty game phases, while their verbal responses suggest otherwise. This leads us to the interpretation that intelligently-designed game mechanics successfully redirected playtesters' negative emotional response to high-uncertainty game phases to thematic and narrative game elements. Post alignment of the team on playtest findings, we present the goal delineation loop where the team reaches a new understanding of the theoretical space, which in turn influences defining new transformed goals for *Outbreak*.

Acknowledgements

Thanks to the SCIPR (Sensing Curiosity in Play and Responding) project and co-PIs Justine Cassell, Jessica Hammer, and L.P. Morency, generously funded by the Heinz Family Foundation. We also thank Anny Fan, Catherine Kildunne, and Eda Zhang for their work iteratively designing and playtesting *Outbreak*.

References

Brooks, A. W. (2014). Get excited: Reappraising pre-performance anxiety as excitement. Journal of Experimental Psychology: General, 143(3), 1144.

Bryant, J., & Miron, D. (2003). Excitation-transfer theory. In J. Bryant, D. Roskos-Ewoldsen, & J. Cantor (Eds.), Communication and emotion: Essays in honor of Dolf Zillmann (pp. 31-59). Mahwah, NJ: Erlbaum.

Culyba, S. (2015). Transformational Games: A Field Guide for Design Leaders. Unpublished manuscript

Flanagan, M., Punjasthikul, S., Seidman, M., Kaufman, G. (2013). Citizen Archivists at Play: Game Design for Gathering Metadata for Cultural Heritage Institutions. *Proceedings of DiGRA 2013, Atlanta, Georgia, August 2013.*

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American psychologist, 56(3), 218.

Frens, J.W. & Hengeveld, B.J. (2013). To make is to grasp. 5th International Congress of International Association of Societies of Design Research (IASDR), 2630 August 2013, Tokyo Tokyo, Japan: Technische Universiteit Eindhoven.

Gordon, G., Breazeal, C., & Engel, S. (2015, March). Can children catch curiosity from a social robot?. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 91-98). ACM.

Gross, J. J., and John, O. P. (2003).Individual differences in two emotion regulation Processes: implications for affect, relationships and well-being. J. Pers.Soc. Psychol. 85, 348–362.

High, R. (2012). The era of cognitive systems: An inside look at IBM Watson and how it works. *IBM Corporation, Redbooks*.

Hofmann, S. G., Heering, S., Sawyer, A. T., & Asnaani, A. (2009). How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal. Behaviour research and therapy, 47(5), 389-394.

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*(2), 125-160.

Kahneman, D., Kahneman, D., & Tversky, A. (2003). Experienced utility and objective happiness: A moment-based approach. The psychology of economic decisions, 1, 187-208.

Kashdan, T. B., & Roberts, J. E. (2006). Affective outcomes in superficial and intimate interactions: Roles of social anxiety and curiosity. *Journal of Research in Personality*, *40*(2), 140-167.

Kaufman, G., & Flanagan, M. (2015). A psychologically “embedded” approach to designing games for prosocial causes. Cyberpsychology: The Journal for Psychosocial Research on Cyberspace, article 5. doi: 10.5817/CP201535.

Kaufman, G., Flanagan, M., & Seidman, M. (2015). Creating stealth game interventions for attitude and behavior change: An “Embedded Design” model. In Proceedings of the Digital Games Research Association (DiGRA) Conference.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*.

Lieberoth, A. (2015). Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture*, *10*(3), 229-248.

Litman, J. A. (2010). Relationships between measures of I-and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences*, *48*(4), 397-402.

Mahdikhani, M., Soheilahamzehloo., Maryamshayestefard., and Mahdikhani, N. (2008). “Student participation in classroom discourse.” vol.5, no. 6, pp. 1422–1431.

Mohammed, S., and Dumville, B. C. (2001). "Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries" in *Journal of Organizational Behavior* 22, no. 2, pp. 89-106.

Nielsen, J. (1993). Iterative userinterface design. Computer, 26(11), 3241.

Paulhus, D. L. (1991). Measurement and control of response bias.

Rocca, C.H., Krishnan, S., Barrett, G. and Wilson, M. (2010). “Measuring pregnancy planning: An assessment of the London Measure of Unplanned Pregnancy among urban, south Indian women.” Demographic research, vol. 23, p.293.

Rozin, P., Guillot, L., Fincher, K., Rozin, A., & Tsukayama, E. (2013). Glad to be sad, and other examples of benign masochism. *Judgment and Decision Making*, *8*(4), 439.

Seidman, M., Flanagan, M., & Kaufman, G. (2015). Failed games: Lessons learned from promising but problematic game prototypes in designing for diversity. In *Proceedings of the Digital Games Research Association (DiGRA) Conference*.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

To, A., Ali, S., Kaufman, G., Hammer, J. (2016). Integrating Curiosity and Uncertainty in Game Design. In Proc. DiGRA/FDG 2016.

To, A., Holmes, J., Fath, E., Zhang, E., Kaufman, G., & Hammer, J. (2017). Modeling and Designing for Key Elements of Curiosity: Risking Failure, Valuing Questions. Proceedings of DiGRA 2017.

To, A., Fath, E., Zhang, E., Ali, S., Kildunne, C., Fan, A., ... & Kaufman, G. (2016). Tandem Transformational Game Design: A Game Design Process Case Study. In Proceedings of the International Academic Conference on Meaningful Play.
Awards : Best Student Non-Digital Game, *Outbreak*, People's Choice Award, *Outbreak*.

Tozour, P. (2015, January 13). The game outcomes project, part 3: Game development factors. Retrieved from http://www.gamasutra.com/blogs/PaulTozour/20150113/233922/The_Game_Outcomes_Project_Part_3_Game_Development_Factors.php#DesignRiskManagement

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*(3681), 269-274.

Zimmerman, J., & Forlizzi, J. (2014). Research through design in HCI. In Ways of Knowing in HCI (pp. 167189). Springer New York